

Purdue University
Purdue e-Pubs

Charleston Library Conference

Creating a Trillion-Field Catalog: Metadata in Google Books

Jon Orwant
Google Books

Anna Fleming
Northwestern University

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Jon Orwant and Anna Fleming, "Creating a Trillion-Field Catalog: Metadata in Google Books" (2010).
Proceedings of the Charleston Library Conference.
<http://dx.doi.org/10.5703/1288284314814>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

answer was “no,” since academic authors direct their work to the reader. Mathematicians, for example, do not feel that a publishers’ brand is needed.

CREATING A TRILLION-FIELD CATALOG: METADATA IN GOOGLE BOOKS

Speaker: Jon Orwant - Engineering Manager, Google Books)

Reported by: Anna Fleming (Northwestern University, Galter Health Sciences Library) <a-fleming@northwestern.edu>

Previously published in Against the Grain, v.23 #6 December 2011/January 2012.

Google Books Engineering Manager Orwant discussed Google’s efforts to create and improve its metadata in order to increase discoverability within its growing collection. He showed examples of the challenges the team faces, including normalizing author names, fun with transliterations, interpreting FRBR rules, and disambiguating multi-volume or serial publications. Clearly, all this sounded familiar to the librarians in the audience. Factoid: as of 18 months ago, the collection included 15 million scanned books in over 480 languages. Orwant also described Google’s new Digital Humanities Awards to researchers interested in linguistic analysis projects. So far, 12 awards to 23 researchers at 15 universities support projects datamining the Google Books corpus. Steven Pinker (author of “How the Mind Works,” “The Stuff of Thought,” and “The Language Instinct”) is a recipient. Stay tuned.

EFFICIENT AND EFFECTIVE FUNDING OF OPEN ACCESS ‘BOOKS’

Speaker: Frances Pinter - Publisher, Bloomsbury Academic; **Sanford G. Thatcher** – Former Director, Penn State University Press; Free-lance Acquiring Editor

Reported by: Ramune K. Kubilius (r-kubilius@northwestern.edu) - Northwestern University, Galter Health Sciences Library

Previously published in Against the Grain, v.23 #6, December 2011/January 2012

Thatcher served as responder for this session during which Pinter provided examples and models of open access book publishing possibilities from her “previous life” with Soros, and her current work at Bloomsbury. With core humanities and social sciences funding, the book format would be brought to the public arena. Calling the European-based eIFL (www.eifl.net) a “whopper of a consortium”, she moved on to academics’ needs and desires: independence, branding, editing, marketing, selling, and, the “mother-in-law” factor (the magnum opus copy for family). Publishers are willing to experiment, see themselves as service providers, not gatekeepers, and co-creators of value. Books are similar to ice cream- “vanilla on HTML”, with a cone (print, e-book, Kindle, etc.), that can be a sundae (the “enhanced e-book”). Arguing that the Open Access STM journal funding model is not sustainable, she envisioned creation of the “International Library Coalition for Open Access Books (ILCOAB)”, to aggregate funds from the